DOCUMENT RESUME

ED 294 316 EA 020 002

AUTHOR Ellett, Chad D.; Garland, JoAnne S.
TITLE Teacher Evaluation Practices in Our Largest School Districts: Are They Measuring Up to "State-of-the-Art" Systems?
PUB DATE Aug 86
NOTE 39p.
PUB TYPE Reports - Research/Technical (143) -- Tests/Evaluation Instruments (160)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Elementary Secondary Education; Evaluation Methods; Evaluation Problems; *Program Effectiveness; *Program Evaluation; School Districts; School Surveys; *State School District Relationship; *Teacher Evaluation
IDENTIFIERS *Large School Districts

ABSTRACT

One hundred of the largest school districts from seven regions were surveyed to scrutinize local teacher evaluation policies and practices, to determine the efficiency of local evaluation instruments and procedures, and to compare local teacher evaluation with state teacher evaluation programs. The Teacher Evaluation Practices Survey (TEPS) was sent to superintendents to gather the data. Generally, local teacher evaluation was inferior to state teacher evaluation. Local teacher evaluation practices had not changed since 1979; local practices exhibited ignorance regarding more recent developments; local practices stressed summative uses of data more frequently than formative uses; few districts reported policies that included teaching performance standards or orientation for evaluators; few districts employed evaluators outside of the school district; local evaluator training was insufficient; local teacher evaluation instruments included criteria that were in opposition to research literature; and many districts employed informal observations of teaching, which opposes state teacher evaluation programs. Thus, the incorporation of recently developed practices of state teacher evaluation programs is deficient in local teacher evaluation. The document includes 11 tables, 14 references, and a copy of the TEPS. (RG)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED294316

Teacher Evaluation Practices in Our Largest School Districts: Are They Measuring Up to "State-of-the-Art" Systems?

Chad D. Ellett and JoAnne S. Garland

College of Education
Louisiana State University

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Chad D. Ellett

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

August, 1986

EA 020 002

Teacher Evaluation Practices in Our Largest School Districts: Are They Measuring Up to "State-of-the-Art" Systems?

Teacher evaluation programs and practices have received renewed emphasis as a means for educational improvement during the 1980's. Earlier approaches to increasing the educational accountability of schools, such as MBO and other results-oriented management models, have seemingly given way to programs and policies targeting improvements in teacher certification, selection and job performance. Holley (1979) described three lines of inquiry that seemed to be converging to mold and shape teacher evaluation practices: (a) teacher effectiveness research, (b) teacher evaluation, and (c) teacher testing (for certification). Each of these lines of inquiry has proliferated during the past few years. In perusing the literature, one can find many syntheses of research and discussions of issues reflecting programmatic and research concerns in these three areas of investigation. In addition, recent reviews of research on effective schools point out the "instructional supervision" function of teacher evaluation practices as an important element in increasing school productivity and learner outcomes (Block, 1983).

An additional line of inquiry that has relevance for understanding teacher evaluation practices and programs is the study of schools as complex social organizations. A recent review and synthesis of teacher evaluation practices within organizational contexts by Darling-Hammond and Wise (1983) points to four minimal conditions for the successful operation of a teacher evaluation system: (a) a shared understanding of teacher evaluation criteria and processes by all "actors," (b) a shared sense by organizational members that the teacher evaluation criteria and processes capture the

most important aspects of teaching, (c) a teacher perception that the evaluation system will enable them to improve performance as well as a principal perception that the evaluation system will improve instructional leadership, and (d) a perception on the part of all actors in the system that the evaluation procedure achieves a balance between control and autonomy. The complexities of implementing sound teacher evaluation practices within the larger instructional management role of school principals have also been explored by Bossert, Dwyer, Rowan, and Lee (1982), Cuban (1984) and more recently by Cuban (1986). Thus, understanding teacher evaluation programs and practices yields a host of complex issues that go beyond more traditional measurement and program implementation concerns.

Perhaps the most visible and controversial uses of teacher evaluation data today center on state-mandated reforms targeting teacher certification, merit pay, career ladder advancement and the like (Street & Ellett, 1986). A large number of states, particularly in the southeast, have initiated the development of complex, large-scale teacher assessment/evaluation procedures in response to the public "press" to increase educational quality. States such as Georgia, Florida, South Carolina, Kentucky, Tennessee, Texas, Virginia, North Carolina, Mississippi and Louisiana have each invested large amounts of human, financial, and technical resources to develop comprehensive, legally defensible and psychometrically sound teacher evaluation systems. These large-scale efforts have generated a variety of teacher observation/evaluation instruments and practices that are generally reflective of the "state-of-the-art" in teacher evaluation. In addition, most of these systems are diagnostic in the sense that they are specifically structured to provide teachers with feedback about perfor-

mance useful for continuing professional development (Ellett, 1986).

Though developed for the purposes of teacher certification, structuring teacher incentive plans (e.g., merit pay in Florida and career ladder in Tennessee) and so on, these large scale evaluation systems may hold promise for the improvement of instruction when used for formative evaluation and instructional supervision purposes (Ellett, 1986). Some of these state-of-the-art systems have been under development and implementation since 1976. The Teacher Performance Assessment Instruments (TPAI) (Capie, Anderson, Johnson & Ellett, 1980), used for the performance-based certification of teachers in Georgia, is a case in point. The TPAI, like the more recently developed large-scale systems such as the Florida Performance Measurement System (FPMS), is a research-based, comprehensive observation and teacher performance evaluation instrument and process that has gained wide acceptability among practitioners in Georgia. The FPMS also has broad-based support in Florida for use in the Florida Beginning Teacher Program. When used in summative evaluation contexts, many important issues pertaining to the content, structure, and use of instruments such as the TPAI and FPMS, have been raised and most have no easy answers (Performance Assessment Systems, Inc., 1985). It is probably fair to say, however, that the extensiveness of development, field testing and revision of large-scale teacher evaluation systems like the TPAI and FPMS place them at the forefront of existing teacher evaluation systems.

Teacher evaluation practices at the local school district level, on the other hand, are not as comprehensive and as carefully constructed as systems like the TPAI and FPMS. Generally, it is not possible for local districts to invest the resources necessary to develop and fine-tune

such comprehensive evaluation instruments and procedures. In addition, local teacher evaluation systems may suffer in uniformity of application because of complex organizational and school context reasons (Darling-Hammond & Wise, 1983); and most can be criticized on methodological and conceptual grounds (Medley, Coker & Soar, 1984).

Moreover, conventional wisdom about effective teaching practices--rather than the extant literature on teacher effectiveness--and local values and norms typically determine the design of various teacher evaluation instruments and procedures used by local school districts. Other factors, such as negotiations with teacher organizations and unions, the values and knowledge level of local school board members, social structures in the school organization (Cuban, 1986), etc., are important determinants of local district teacher evaluation criteria and program implementation procedures.

Purpose

Holley (1979) previously reported the results of a national survey which documented the wide variety of evaluation criteria and practices in the 100 largest school districts' teacher evaluation programs. The purpose of this study was to extend and update Holley's initial inquiry in view of the recent development of large-scale teacher assessment programs. In addition, an analysis of the quality of evaluation instruments and methodologies currently being used in these districts was completed. Of particular interest was how local school district teacher evaluation systems measure up to critical features of large-scale, state-of-the-art teacher evaluation systems developed in response to state mandates during the past 8 to 10 years.

Methodology

Data Sources

A list of the 100 largest school districts in the United States was obtained from the Curriculum Information Center School Directory (1982-83). The sizes of the school districts, based on student enrollment ranged from approximately 33,000 to 932,000 students. The directory also provided the names and addresses of the superintendents in each school district; and Patterson's American Education (Moody, 1985) was used to update any personnel changes that might have occurred since the directory's publication.

Instrumentation

Data for the study were collected with the Teacher Evaluation Practices Survey (TEPS). Items on the TEPS were designed to reflect critical features of state-of-the-art programs. The TEPS consisted of four major components: (a) purposes, (b) policies, (c) practices, and (d) opinions. The first three components were operationalized by 4, 7, and 13 items respectively, and each of these items was further defined by separate sub-items. The fourth component consisted of five items designed to assess respondents' summary opinions about the teacher evaluation practices in their respective school districts. A copy of the TEPS can be found in Appendix A.

Data Collection Procedures

In May of 1985, a packet which included a TEPS and a letter explaining the survey and requesting copies of relevant documents pertaining to teacher evaluation practices was mailed to each district superintendent. Superintendents were asked to either complete the survey personally or

to have a member of their staff complete it. Six weeks later, follow-up letters were mailed along with self-addressed, stamped envelopes to all non-respondents. This second request resulted in an overall return rate of 70 percent. In an effort to increase the return rate, a third follow-up was conducted in the late fall of 1985. Consequently, the results of this study reflect information supplied by 80 of the 100 largest school districts in the United States.

Data Analysis

Descriptive statistics for each survey item were generated. In an attempt to quantify survey data for analysis purposes, a "Quality Index" (QI) score for each TEPS was computed by assigning points to each TEPS item to reflect preferred, critical features of state-of-the-art teacher evaluation practices and programs. For example, a respondent could check ten items representing critical features of the districts' evaluator training program. One point was assigned for each item checked in this section of the survey.

An additional survey component which reflected respondents' summary opinions about five effects of the district teacher evaluation program was also constructed. Intercorrelations between the TEPS QI and the respondents' summary opinions were computed to examine the relationship between the quality of teacher evaluation program components and the effectiveness of the various programs as perceived by TEPS respondents.

Qualitative analyses of supporting documentation submitted by respondents (evaluation instruments, program procedures, etc.) were also completed. The intent of these analyses was to better understand the degree to which these documents corroborated responses to the survey.

Results

Results of the study are divided into two sections. The first section details the quantitative analyses of the survey data, and the second section provides a summary of the qualitative analyses of the supporting documentation.

Quantitative Analyses - Survey Data

Demographics. The eighty school districts supplying survey data were distributed among seven geographic regions in the United States. Nine or 11.25 percent of the participants were located in 6 mideast states, 7 (8.75%) in 4 Great Lake states, 5 (6.25%) in 4 Plains states, 34 (42.5%) in 11 southeast states, 12 (15.0%) in 4 southwest states, 3 (3.75%) in 2 Rocky Mountain states and 10 (12.5%) in 4 far west states. It should be noted that only one New England region school district was in the survey population, and it was among the non-respondents.

Table 1 summarizes demographic information about the school districts participating in the study. The size of the schools based on the number of students enrolled ranged from 33,000 to 932,000. The mean number of schools and teachers is provided for each school-size category and ranged from 65 to 906 and from 2,361 to 42,751, respectively.

Purposes. This section of the TEPS contained questions concerning the purposes and uses of the district teacher evaluation system. First, respondents were asked to rank order four commonly stated purposes of teacher evaluation in terms of their importance considering policies and practices within their own school district. Percentages of respondents assigning ranks from 4 (most important) to 1 (least important) to each

Table 1

Total Response by Size of School District

Number of Respondents	Number of students	Mean number of schools	Mean number of teachers
33	33,000 - 49,999	65	2,361
23	50,000 - 74,999	91	3,215
11	75,000 - 99,999	130	5,185
9	100,000 - 199,999	184	7,163
2	200,000 - 499,999	254	12,089
2	500,000 - 932,000	906	42,751

purpose are shown in Table 2. On the average, professional development for teachers (x=2.70) was regarded as the most important purpose of teacher evaluation; whereas, instructional leadership for administrators (x=2.39) was ranked as least important. Percentages of respondents ranking the purposes Most Important ranged from 36.8 (Professional development for teachers) to 14.5 (Personnel decisions).

Table 2

Ranking of Teacher Evaluation Purposes
(n=76)

Purpose	Rank 4	3	2	1	Mean Rank
Accountability	27.6	25.0	26.3	21.1	2.59
Personnel decisions	14.5	38.2	27.6	19.7	2.47
Professional development for teachers	36.8	18.4	22.4	22.4	2.70
Instructional leadership for administrators	26.3	19.7	21.1	32.9	2.39

Next, questions regarding the utilization and treatment of evaluation results were posed. Table 3 provides the distribution of responses to the question, "How are evaluation results used in your school system?" The items receiving the largest percentages of responses (90% and over) were: "Development of remediation plans for teachers with identified deficiencies" and "Teacher dismissal." Clearly, the two most popular uses involve either helping deficient teachers or dismissing them. Less than a fourth of the respondents indicated that evaluation results were used for certification and merit pay decisions, and only one half said that results were used in the promotion/compensation of teachers.

Table 3

Use of evaluation results

Use	Percent
Certification decisions	21.3
Development of remediation plans for teachers with identified deficiencies	92.5
Merit pay decisions	23.8
Placing teachers on probation	58.8
Renewal of teachers' contracts	73.8
Teacher dismissal	90.0
Teacher promotion/compensation	50.0
Tenure decisions	71.3

According to the TEPS results, official summaries of evaluation data are compiled and reported for all teachers by 35 (45.5%) of the school districts represented in this study, for each school by 23 (32.4%) and for subgroups of teachers (e.g., elementary; secondary) by 17 (24.6%). Somewhat surprisingly, TEPS respondents reported that evaluation data

are held confidential at the school site in only 52 (65.0%) of the participating school districts.

Policy. This section of the TEPS contained questions concerning teacher evaluation policies in the school district. All but three respondents (97.5%) indicated that there was a formal written teacher evaluation policy currently in effect in their school system? Table 4 provides the percentage of responses to the question, "Which of the following groups formulated the teacher evaluation policy?" The item receiving the largest percentage of responses (87.5%) was Local School Board. The groups specified most often in Other were teacher unions (23.8%) and various assocations representing administrators and teachers (21.3%).

Table 4

Groups Formulating Teacher Evaluation Policy

Group	Percent
State Legislature	47.5
State Department of Education	36.3
State Board of Education	35.0
Local School Board	87.5
Other	45.0

Respondents were then asked to identify groups which had input in the development of the teacher evaluation policy. Table 5 indicates that teachers, administrators and central office staff had input in over 90 percent of the school districts; whereas business leaders and students were involved in teacher evaluation policy development in less than ten percent of the school districts.

Table 5

Groups Involved in the Development of Teacher Evaluation Policy

Group	Percent
Business leaders	8.8
State Department of Education staff	25.0
Central office staff	97.5
School site administrators	97.5
Teachers	91.3
Teacher Union representatives	65.0
Teacher Organization representatives	37.5
Parents	18.8
Students	7.5

A relatively even distribution of responses was received for the question, "How long has the current policy been in effect?" Approximately one fourth of the school districts fell into each of the following categories: more than 10 years (26.3%), from 5 to 10 years (25.0%), from 3 to 5 years (21.3%), and less than 3 years (27.5%).

The TEPS results indicated that the most popular medium for informing teachers of the teacher evaluation policy is the faculty meeting (87.5%) followed by the teacher handbook (72.5%), the inservice meeting (68.8%), and the teacher contract (48.8%). Mentioned Other (13.8%) were individual conferences, videotapes, syllabi and copies of the evaluation policy.

Ninety-five percent (76) of the respondents indicated that their teacher evaluation policy was comprehensive (i.e., precisely written with clearly worded details of procedures to be followed). These respondents were asked to check details which were specified in their policy. Table 6 shows that four basic details are included in the policies of almost

all of the school districts represented in the study. However, only 70.7 percent of the respondents indicated that standards for acceptable teaching were included in their policy. Moreover, only 61.3 percent indicated that their policy provided for orientation of evaluators to the evaluation instrument.

Table 6

Details Specified in Teacher Evaluation Policies
(n=75)

Detail	Percent
Who will be evaluated	98.7
How many times evaluation(s) will take place	100.0
Who will conduct evaluation(s)	97.3
Number of evaluations required for different kinds of teachers	92.0
Standards for acceptable teaching	70.7
Orientation of evaluators to the evaluation instrument	61.3

The last question in this section was "Does the policy stipulate systematic review and revision of teacher evaluation practices?" Of the 77 respondents answering this item, only 31 (40.3%) marked Yes.

Practices. This section of the TEPS contained questions concerning the teacher evaluation practices followed in each school district. One objective of this section was to determine who conducted official evaluations and more critically, the extent of their training in teacher evaluation procedures.

Table 7 shows that building level administrators are official evaluators of teaching in all of the school districts represented in this study. At the school district level, however, supervisors of instruction are allowed to conduct official evaluations in 57.5% (46) of the school districts. Those least often indicated as being official evaluators were students (2.5%), school board members (3.8%), research and development personnel (7.6%) and teachers (10.0%).

Table 7

Official Evaluators

Official Evaluators	Percent
District Level:	
School Board Members	3.8
Superintendent	36.3
Asst./Associate Superintendent	40.0
Supervisors of Instruction	57.5
Personnel Officers	18.8
Research & Development Personnel	7.5
Building Level:	
Principal	100.0
Assistant Principals	92.5
Teachers	10.0
Students	2.5

Fifty three (66.3%) of the respondents indicated that comprehensive evaluation training (two or more days) was required of evaluators. These respondents answered several questions regarding the extent of evaluator training. Responses to the first question dealing with the content of the training program are shown in Table 8.

The response to "Who conducts the training of evaluators?" fell into three categories: a) staff development/personnel management (73.6%), b) state departments of education (7.5%), and c) professional consultants (13.2%). The length of training programs varied from a few hours to several days. A breakdown of the responses to "How many hours of evaluation training are required?" follows: an unspecified amount (17.0%), less than 8 hours (15.1%), from 8 to 15 hours (9.4%), from 16 to 23 hours (9.4%), from 24 to 35 hours (28.3%), from 36 to 47 hours (3.8%), 48 hours or more (5.7%).

Table 8

Content of Evaluator Training Program
(n=53)

Content	Percent
Information about:	
Procedures to be followed	98.1
Observation techniques	96.2
Evaluation instrument items/criteria	98.1
Conducting conferences with teachers	98.1
Examination of lesson plans	60.4
Validity and reliability issues	56.6
Orienting teachers to evaluation procedures	94.3
Legal issues in evaluation	73.6
Practice evaluations using videotaped teaching situations	69.8
Observations of teachers in actual classroom settings	54.7

Importantly, only twenty respondents (37.7%) indicated that participants in training were required to pass a "proficiency test" (i.e., tested to demonstrate that they have sufficient knowledge and reliability to

use the evaluation instrument appropriately). A slightly higher number, 30 (56.6%), indicated that certified teacher evaluators were required to participate in "update" training activities after initial training.

Another objective of this section of the survey was to gain information about teacher evaluation procedures. All but seven school districts (91.3%) reported that the procedures to be followed in teacher evaluation were clearly written and detailed.

Table 9 shows that both formal and informal observations are methods of evaluation most often included in teacher evaluation plans. All but one of the respondents disclosed that direct, systematic observation of teaching was a component of their school district's teacher evaluation plan; 85 percent also checked informal observation of teachers. Low percentages of responses appeared in several categories indicating that these methods are not widely used. These were: student achievement data (17.5%), peer ratings of teacher's performance (3.8%), paper and pencil examinations (10.1%), and student ratings of teacher's performance (3.8%).

Table 9

Methods of Evaluation

Methods	Percent
Direct, Systematic Observation of Teaching	98.8
Informal Observation of Teachers	85.0
Peer Ratings of Teacher's Performance	11.3
Student Ratings of Teacher's Performance	3.8
Student Achievement Data	17.5
Paper and Pencil Examinations	10.0
Teacher Self Evaluation	31.3
Other	8.8

Those respondents indicating that systematic observation of teaching was a required evaluation procedure were asked to further describe features of the observation. Table 10 contains a summary of percentages of respondents checking various details included in the observation process.

Table 10

Systematic Observation Details

Detail	Percent
Standardized observation form	72.2
Specified length of each observation	67.1
Specific schedule or timelines for completing observations	79.7
Pre-observation conference	58.2
Examination of lesson plans prior to observation	40.5
Pre-observation interview	30.4
Post-observation interview	45.6
Post-observation conference	88.6

When asked how many official observations were required, a wide range of answers was given by TEPS respondents (n=76). For tenured teachers, the response ranged from none required (1.3%) to once every five years (3.9%) to as many as six times a year (1.3%). The most common numbers of required observations were one or two each year for tenured teachers (56.6%) and two or three per year for non-tenured teachers (68.0%). For teachers with identified deficiencies, however, the majority of responses (74.3%) revealed that no specified number of observations was required.

Only 21 respondents (26.6% of 79) answered Yes to the final question regarding teacher observation, "Are there written rules for combining the

results of multiple observations?"

A summary of the responses to the last TEPS questions which addressed various issues and practices follows. In 75 school districts (93.8%), teachers are required to sign evaluation forms. In 62 school districts (77.5%), a formal grievance procedure exists for teachers to contest negative evaluations; and in 73 school districts (91.3%), a formal grievance procedure exists for teachers to contest violations of procedures. However, only 19 school districts (23.8%) specify consequences for evaluators who violate procedures. A remediation plan for teachers who receive negative evaluations is included in the evaluation process in 68 school districts (85%); furthermore, in 59 school districts (73.8%), the evaluator is obligated to develop a professional improvement plan for the teacher if teaching deficiencies are detected.

Opinions. This section of the survey instrument contained questions concerning opinions about the teacher evaluation system used in each school district. The first question was "How adequate is the teacher evaluation system in accomplishing the purposes for which it was designed?" Twenty respondents (32.5%) checked More than adequate, 44 ((55.0%) checked Adequate, and 6 (7.5%) checked Less than adequate. Four (5.0%) did not respond.

Next, respondents were asked to indicate to what extent the teacher evaluation system had impacted the manner in which evaluators viewed classroom teaching. The answers were: Much impact, 41.3%; Some impact, 50.0%; and Little impact, 5.0%. Three (3.8%) did not respond.

A third question dealt with the extent to which the teacher evaluation system had served to communicate clearer expectations of teaching to

teachers. Half of the respondents (50.6%) indicated that much impact had been made in this area, and 40.0 percent reported that some impact had been made; whereas 7.5% of the respondents felt that little impact had occurred.

Less than a third of the respondents (30.0%) replied Much improvement to the question, "To what extent has the teacher evaluation system served to improve instruction?" About a half (53.8%) marked Some improvement, 7.5 percent marked Little improvement and 8.8 percent did not answer this question.

The final question dealt with the extent to which the teacher evaluation system had led to better/more systematic supervision of teaching in the respondent's school district. Again, over a half of the respondents answered positively. Much impact was checked by 42 respondents (52.5%), 30 (37.5%) checked Some impact and only 4 (5.0%) checked Little impact.

Intercorrelations

Table 11 presents a summary of intercorrelations between TEPS respondent opinion variables and the TEPS "Quality Index." The QI was computed by aggregating respondent scores for survey items checked that were considered to be desirable characteristics of a state-of-the-art teacher evaluation instrument. QI scores ranged from 22 to 88 with a mean of 55.42. QI scores were, for the most part, equally distributed throughout the possible score range.

Table 11

Intercorrelations Between TEPS Respondent Opinion Variables and the TEPS QI Score

Opinion Variable	TEPS QI Score
	r
Adequacy	.51
Impact	.52
Expectations	.46
Instructional Improvement	.47
Supervision	.47

All of the correlations in Table 11 are highly statistically significant ($p < .0001$), positive in direction, and in the moderate range. Consistent, positive relationships existed between the teacher evaluation quality index score and respondents' opinions about the extent to which they perceived their teacher evaluation system as: (a) adequate for accomplishing the purposes for which it was designed, (b) impacting the manner in which district evaluators viewed classroom teaching, (c) communicating expectations about classroom teaching, (d) serving to improve instruction in the district, and (e) leading to more and/or better systematic supervision of teaching.

Qualitative Analyses - Supporting Documentation

Qualitative analyses were completed using evaluation policy manuals and instruments from the 30 school districts that furnished the requested documentation (25 submitted copies of both policy description and evaluation instruments, 3 submitted only the instrument, and 2 submitted only a policy description). The purpose of these analyses was to examine (a) the philosophical basis and goals of the evaluation program, (b) the methodological/

measurement framework of the evaluation instruments, (c) the specific behavioral or personal characteristics evaluated, and (d) the extent to which the assessment procedures reflected due process and professional development concerns.

Philosophy and Goals: Of the 27 teacher evaluation policy descriptions reviewed, 25 included a statement of philosophy, purpose or goals as a basis to the evaluation plan. The most prevalent theme throughout these descriptions was the belief that teacher evaluation results can improve instruction which, in turn, results in increased pupil learning. Consequently, the most popularly stated purposes and goals of the teacher evaluation plans centered on some aspect of instructional improvement. Some statements of program philosophy assumed that teacher evaluation would foster not only professional development but also increased job satisfaction.

Methodological/Measurement Framework: Twenty-nine instruments were scrutinized to determine the: (a) criteria used to assess teaching competence, and (b) the methodology used to evaluate teacher performance. These instruments varied considerably in the number and kinds of criteria used to evaluate teachers and the degree to which these criteria were operationalized in terms of observable teacher behaviors. The vast majority of the instruments categorized instructional criteria into broad areas of teacher responsibility such as Preparation and Planning, Classroom Organization and Management, and Professional Growth and Attitudes. The number of categories varied from 2 to 23. The number of performance criteria or teaching "competencies" assessed varied from 10 to 65. Ten of the instruments included subsets of behavioral indicators

to define criteria or more detailed explanations of the criteria to be judged. Interestingly, only three of these ten were structured so that indicators could be assigned numerical values.

These instruments also differed considerably in the manner in which summative decisions of overall "acceptability" of performance were made. All but three observation forms required a singular summative judgment by the evaluator given whatever cumulative evidence was at hand. In 20 of the instruments analyzed, a categorical summative decision is made about teachers' competencies without referring to any explicit decision or data aggregation rules. Seventeen instruments provided for decisions such as Satisfactory, Needs Improvement and Unsatis- factory. Other typical decision categories were: Effective, Not Effective, Superior, Good, Commendable.

Teaching Behaviors Assessed. Most of the instruments seemed to encompass at least some of the teacher behaviors supported by research on teacher effectiveness (e.g., individualization, classroom management, time usage, clarity, etc.). Few instruments, however, appeared to have been designed to specifically reflect the extant research literature on teacher effectiveness. In addition, most instruments included items that focused on personal characteristics of teachers such as appearance, punctuality, relations with colleagues, professionalism, professional attitudes, and other non-instructional variables.

Due Process and Professional Development. For the most part, the teacher evaluation policy descriptions revealed a concern for due process and fairness for teachers. Twenty-four descriptions contained detailed procedures including: who was to be evaluated, persons designated as

official evaluators, number of required evaluations, and grievance procedures. Twenty-one descriptions specified teacher evaluation time-lines and 12 included definitions of terms.

The most common approach to professional development of teachers among the school districts was the "contract approach." Policies from 19 of the school districts specified professional development programs wherein the teacher is expected to write individual accountability plans and performance objectives at the beginning of the school year. During the course of the year, periodic observations are made to assess the teacher's progress toward meeting the objective(s). At the end of the school year, a summative decision based on a combination of informal and formal observations and the teacher's self evaluation report is made as to whether or not the teacher satisfactorily accomplished the objective(s).

A total of 26 districts (including the 19 districts using the contract approach) provided for teacher improvement through supervision, inservice, college courses, and/or other developmental resources. These services were restricted for use to only those teachers whose teaching performances were judged as below standard. Provisions for the continuing professional development of teachers who perform satisfactorily were not routinely made. Twenty-one district policies stipulated the use of evaluation data to justify dismissal or other negative personnel actions.

Discussion and Conclusions

The purpose of this study was to examine the nature of teacher evaluation policies and practices in the 100 largest school districts in the United States. Of particular interest was the extent to which

local school district practices and procedures reflect important issues, concerns and critical features identified through the recent development of large-scale state-of-the-art teacher evaluation systems being implemented in a variety of states. The assumption was made that large-scale systems, more recently developed in response to state mandates for educational reform and increased teacher quality, are the best examples of the state-of-the-art in teacher evaluation.

Survey data from 80 of the 100 largest school districts were analyzed and analyses of evaluation instruments, program descriptions and policy manuals submitted by districts were completed. These analyses yielded a variety of interesting results that can be interpreted in view of more recent developments and practices emanating from large-scale teacher evaluation programs such as those under development or being implemented in a number of states (e.g., Georgia, Florida, Tennessee, Virginia, Mississippi, Louisiana, Texas, etc.).

For the most part, it appears that teacher evaluation practices and policies of local school districts are lagging behind in incorporating important elements of teacher evaluation practices being identified through larger state-level efforts. This is certainly understandable given the amount of human, technical, and financial resources required to develop psychometrically and procedurally sound systems derived from the emerging research base on effective teaching. However, the issues identified in reviewing the data in this study point to several concerns about teacher evaluation instruments, practices and procedures currently in use at the local school district level.

The results of the TEPS analyses suggest that approaches being

taken to teacher evaluation and the philosophical basis on which local districts establish such systems have not significantly changed since the data reported by Holley (1979). This overall view suggests that local district teacher evaluation systems are lagging behind in terms of being informed and influenced by more recent developments in the field.

An overwhelming percentage of respondents, for example, stated that the primary purpose and use of teacher evaluation data in their districts are teacher dismissal and remediation. Qualitative analyses of supporting documentation, however, suggested that few districts seemed concerned with broader applications such as staff development for school principals, continuous professional development for all teachers, increasing school effectiveness/productivity, etc. Far more emphasis in these large school districts is being placed on summative rather than formative uses of teacher evaluation data. Additionally, under-utilization of evaluation data for formative evaluation purposes seems evident, since only 25 to 50 percent of the respondents reported that data are being examined and used for these purposes.

The policy base of many local district teacher evaluation systems also seems somewhat deficient when compared to critical features of large-scale assessment systems. Of particular concern in this regard are establishing standards for teaching performance and implementing comprehensive training programs to orient evaluators to the teacher evaluation system. Of the respondents surveyed with the TEPS, only 29.3 percent stated that policies included performance standards. This was corroborated by the qualitative analyses of documentation submitted

by districts. Almost 60.0 percent stated that district policies did not stipulate systematic review/revision of their teacher evaluation system, and, surprisingly, only 38.7 percent stated that orientation for evaluators was included in district policies.

The kinds of local district evaluators specified by respondents to the TEPS were those that traditionally have been established by school boards as evaluators (principals, assistant principals and supervisors of instruction). Thus, local evaluation practices still focus on building level concerns, and few allow for the use of evaluators external to the school district or for the inclusion of colleague or peer teachers as data collectors.

Many large-scale teacher evaluation systems have moved in this direction to control the potential adverse effects of evaluation context variables on the evaluation process and to increase the reliability and credibility of evaluation data (Capie, 1985). For example, it is known from a recent study of large-scale teacher evaluation systems that performance scores are considerably inflated when data are collected under conditions that have potentially serious consequences for teachers (and evaluators such as school principals). A recent analysis of data from 7,400 observations of teaching to make merit pay decisions with the Teacher Assessment and Development System (TADS) in the Dade County, Florida schools is a case in point. Florida's merit pay effort required only two observations--one by the building principal and a second by another principal.[1] When these TADS scores, obtained under merit pay conditions, were compared to performance data collected under research and development conditions, considerable score inflation was observed

(Performance Assessment Systems, 1985). Similar results have been reported from preliminary analyses of teacher performance data using the Texas Teacher Appraisal System to meet requirements of the Texas Career Ladder Law. Thus, performance evaluation is not "context free" when important personnel decisions are pending and local school districts should perhaps design procedures to accommodate this finding. Issues of practicality, admittedly, often vitiate the possibility of including external evaluators in teacher evaluation systems at the district level. But, the methodological and potential instructional improvement benefits may eventually outweigh this concern.

Interestingly, 66.3 percent of the TEPS respondents indicated that comprehensive training was provided for evaluators. However, only 37.7 percent indicated that evaluators received at least 24 hours of training and slightly more than half (54.7%) indicated that training included systematic observations of teachers in actual classroom settings. In addition, approximately 63.0 percent stated that trainees did not have to meet a reliability standard in the use of the evalution instrument during training. In large-scale teacher evaluation programs, the evaluator training component is a critical element in establishing sufficient observer reliability and observations of teachers in real classroom settings seems a must. Approximately 30 percent indicated that neither videotape nor practice in actual classrooms was included in the training of evaluators.

Similarly, only 56.6 percent of TEPS respondents stated that instrument validity and reliability issues were discussed in evaluator training. The qualitative analysis showed that many school districts' teacher evaluation

instruments depend totally upon personal judgments of evaluators who typically receive little or no systematic training. Moreover, a wide variety of assessment criteria was evident in the various instruments examined. However, relatively few criteria were at a level of behavioral specificity necessary to train observers to make reliable assessments of teacher performance. This seems a critical concern given the potential seriousness of negative evaluation decisions.

While the reliability of evaluators is an important methodological and training concern for local districts, the validity of currently used evaluation instruments is also highly important. Many of the instruments reviewed in this study encompassed some elements of teaching that are reasonably well-documented in the teacher effectiveness literature. However, most contained a variety of teaching behaviors and scoring criteria that were not consistent with or derived from the research literature. Knowing "what" should be observed/evaluated and "why" may be equally or more important to school principals and other district-level evaluators than accurate observation, data interpretation and instrument scoring. This seems particularly the case from the formative evaluation perspective of using performance assessment data to improve teaching.

Procedurally, TEPS respondents stated that both informal (85.0%) and formal (98.8%) observations of teaching were part of their evaluation system. This finding is consistent with traditional building level practices, but is somewhat at odds with the high degree of formality reflected in more recent large-scale programs. Informal, unstandardized procedures can be criticized because of the equity and fairness in application

issues they raise among teachers. In addition, unstandardized procedures, such as variations in the length and number of observations have many implications for aggregating data to make summative evaluation decisions.

Only 26.6 percent of TEPS respondents stated that there were specific written rules for combining the results of multiple observations. Interestingly, the qualitative analysis of policy and procedures documents submitted for review showed only one clear statement and detailed explanation of such procedures for combining data. This procedure had been jointly developed and endorsed by representatives of the school system administration and the local teachers' union. Most large-scale teacher assessment systems have clear data combining and decision-making rules derived from multiple evaluations. This is a critical concern that should be addressed in designing district level teacher evaluation systems.

Intercorrelations between the TEPS Quality Index and the respondents' perceived quality of local district teacher evaluation programs showed moderate, positive relationships. Thus, the quality of district level teacher evaluation practices is associated with perceived influence of these practices on school outcomes and instructional improvement. However, when TEPS QI scores were compared to the actual quality of evaluation instruments and documentation submitted for review, little or no relationship was apparent. This finding may suggest that survey respondents had a tendency to present a more favorable view of the quality of their teacher evaluation programs, policies and practices than actually existed.

In conclusion, it is apparent that translation of state-of-the-art practices in teacher evaluation from large-scale assessment applications

to local school district uses seems sorely lacking. The results of this study suggest vast differences between the products of newer, large-scale research and development efforts in teacher evaluation during the past 8 to 10 years and the influence of such developments on adaption/adoption practices of local school districts. Little seems changed since Holley's (1979) survey relative to the quality of teacher evaluation practices in the 100 largest school districts. Establishing mechanisms to communicate findings from large-scale teacher assessment programs to better impact local school district evaluation practices seems a critical concern to the field of teacher evaluation.

Notes

1. The Florida Master Teacher Program law was repealed in June, 1986 by the Florida State Legislature; thus, merit pay evaluations are no longer occurring.

References

Block, A. W. (Ed.). (1983). Effective schools: A summary of research. Arlington, VA: Educational Research Service.

Bossert, S. T., Dwyer, D. C., Rowan, B., & Lee, G. V. (1982). The instructional management role of the principal. Educational Administration Quarterly, 18(3), 34-64.

Capie, W. (Fall, 1985). Coming of age: Systematic performance evaluation. Educational Measurement: Issues & Practices, 4(3), 13-16.

Capie, W., Anderson, S. J., Johnson, C. E. & Ellett, C. D. (1980). Teacher Performance Assessment Instruments. Georgia Department of Education, Teacher Assessment Project, College of Education, University of Georgia, Athens, GA.

Cuban, L. (September, 1986). Persistent instruction: Another look at constancy in the classroom. Phi Delta Kappan, 68(1), 7-11.

Cuban, L. (May, 1984). Transforming the frog into a prince: Effective schools research, policy and practice at the district level. Harvard Educational Review, 54(2), 129-151.

Curriculum Information Center's School Directory. (1982/83). Westport, CT: Market Data Retrieval.

Darling-Hammond, L. & Wise, A (Fall, 1983). Teacher evaluation in the organizational context: A review of the literature. Review of Educational Research, 53(3), 285-328.

Ellett, C. D. (Fall, 1986). Emerging teacher performance assessment practices: Implications for the instructional supervision role of school principals. In W. Greenfield (Ed.), Instructional leadership: Concepts and controversies. Boston: Allyn & Bacon.

Holley, F. (1979). Teacher testing, teacher evaluation and research on teaching: Converging fields? Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Medley, D. M., Coker, H. & Soar, R. S. (1984). Measurement-based evaluation of teacher performance: An empirical approach. New York: Longman.

Moody, D. (Ed.). (1985). Patterson's American Education. Illinois: Educational Directories.

Performance Assessment Systems (October, 1985). The use of the Teacher Assessment and Development System (TADS) - MPT Form instrument to meet requirements of Florida's Associate Master Teacher Program in 1984-85: Research report for external reviewers. Miami: Dade County Schools.

Street, S. & Ellett, C. D. (1986). An analysis of policy-based responses educational reforms for teachers: Performance-based certification, merit pay and career ladders. Paper presented at the annual meeting of the Southwestern Educational Research Association, Houston, TX.

Appendix A

Teacher Evaluation Practices Survey

TEACHER EVALUATION PRACTICES SURVEY

INSTRUCTIONS

Please supply the background information requested below and respond to each item on the attached Teacher Evaluation Practices Survey Form. Be certain to note any special instructions (e.g., "Check all that apply."). If you do not know the correct answer to a particular question, please ask an appropriate colleague in your school district to supply you with needed information.

1. Name/Address of School District _____

   _____

   _____

   _____

2. Size of School District

   a. Number of students _____

   b. Number of schools _____

   c. Number of teachers _____

3. Position of Person(s) Completing the Attached Survey Form (e.g., Superintendent, Director of Personnel, etc.)

   _____

   _____

   _____

TEACHER EVALUATION PRACTICES SURVEY

I. PURPOSES

This section contains questions concerning the purposes and uses of the teacher evaluation system in your school district.

1. Listed below are commonly stated purposes of teacher evaluation. Considering policies and practices in your school system, rank these goals as to their importance from 4 (most important) to 1 (least important).

Accountability	
Personnel decisions (e.g., promotion, dismissal)	___
Professional development for teachers	___
Instructional leadership for administrators	___
	___

2. How are evaluation results used in your school system?

   As documentation or support for: (Check all that apply.)

Certification decisions	
Development of remediation plans for teachers with identified deficiencies	___
Merit pay decisions	___
Placing teachers on probation	___
Renewal of teachers' contracts	___
Teacher dismissal	___
Teacher promotion/compensation	___
Tenure decisions	___
	___

3. Are official summaries of evaluation data compiled and reported by your school district?

For all teachers	Yes ___	No ___
For each school	Yes ___	No ___
Subgroups of teachers (e.g., elementary; secondary)	Yes ___	No ___

4. Are evaluation data held confidential at the school site? Yes ___ No ___

II. POLICY

This section contains questions concerning teacher evaluation policy in your school district.

1. Is there a formal written teacher evaluation policy currently in effect in your school system? Yes ___ No ___

2. Which of the following groups formulated the teacher evaluation policy? (Check all that apply.)

State Legislature	
State Department of Education	___
State Board of Education	___
Local School Board	___
Other (Please specify)_____	___

3. Which of the following groups had input in the development of the teacher evaluation policy? (Check all that apply.)

Business leaders	___
State Department of Education staff	___
Central office staff	___
School site administrators	___
Teachers	___
Teacher Union representatives	___
Teacher Organization representatives	___
Parents	___
Students	___

4. How long has the current policy been in effect?

More than 10 years	___
From 5 to 10 years	___
From 3 to 5 years	___
Less than 3 years	___

5. How are teachers informed of the policy? (Check all that apply.)

Contracts	___
Faculty Meetings	___
Inservice Meetings	___
Teacher Handbooks	___
Other (Please specify)	_____

6. Is the policy comprehensive? (i.e., precisely written with clearly worded details of procedures to be followed) *Yes ___ No ___

*If yes, which of the following details are specified in the policy? (Check all that apply.)

6.1 Who will be evaluated	___
6.2 How many times evaluation(s) will take place	___
6.3 Who will conduct evaluation(s)	___
6.4 Number of evaluations required for different kinds of teachers (i.e., beginning teachers v.s experienced/tenured teachers)	___
6.5 Standards for acceptable teaching	___
6.6 Orientation of evaluators to the evaluation instrument	___

7. Does the policy stipulate systematic review and revision of teacher evaluation practices? Yes ___ No ___

PLEASE CONTINUE ON PAGE 3

III. PRACTICES

This section contains questions concerning the teacher evaluation practices that are followed in your school district.

1. Who is allowed to conduct official evaluations of teaching in your school district? (Check all that apply.)

District Level:
- School Board Members ___
- Superintendent ___
- Asst./Associate Superintendent ___
- Supervisors of Instruction ___
- Personnel Officers ___
- Research & Development Personnel ___

Building Level:
- Principal ___
- Assistant Principals ___
- Teachers ___
- Students ___

2. Is comprehensive evaluation training required of evaluators? (e.g., training sessions of two or more days) *Yes __ No __

*(If yes, please answer the following questions.)

2.1 Which of the following is included in the evaluator training program?: (Check all that apply.)

Information about:
- Procedures to be followed ___
- Observation techniques ___
- Evaluation instrument items/criteria ___
- Conducting conferences with teachers ___
- Examination of lesson plans ___
- Validity and reliability issues ___
- Orienting teachers to evaluation procedures ___
- Legal issues in evaluation ___

Practice evaluations using videotaped teaching situations ___

Observations of teachers in actual classroom settings ___

2.2 Who conducts the training of evaluators? _____

2.3 How many hours of evaluator training are required? _____

2.4 Are participants in training required to pass a proficiency test? (i.e., tested to demonstrate that they have sufficient knowledge and reliability to use the evaluation instrument appropriately) Yes __ No __

2.5 Are certified teacher evaluators required to participate in any "update" training activities after initial training? Yes __ No __

3. Are the procedures to be followed in teacher evaluation clearly written in detail? Yes __ No __

4. Which of the following methods of evaluation are included in your district's teacher evaluation plan? (Check all that apply.)

Direct, Systematic Observation of Teaching ___
Informal Observation of Teachers (i.e., occasional unannounced visits) ___
Peer Ratings of Teacher's Performance ___
Student Ratings of Teacher's Performance ___
Student Achievement Data ___
Paper and Pencil Examinations ___
Teacher Self Evaluation ___
Other (Please specify) _____

5. If your teacher evaluation system requires systematic classroom observation, which of the following are included? (Check all that apply.)

Standardized observation form ___
Specified length of each observation (e.g., minimum of 30 minutes) ___
Specific schedule or timelines for completing observations (e.g., one before Jan. 1, etc.) ___
Pre-observation conference (i.e., lengthy, official procedure to prepare the teacher for observation) ___
Examination of lesson plans prior to observation ___
Pre-observation interview (i.e., brief questioning of the teacher immediately prior to observation) ___
Post-observation interview (i.e., brief questioning of the teacher upon exiting the classroom for clarification purposes) ___
Post-observation conference (i.e., lengthy, detailed discussion of teacher's strengths & weaknesses) ___

6. How many official observations are required? (e.g., once each year; once every three years)

For tenured teachers _____
For non-tenured teachers _____
For teachers with identified deficiencies _____

7. Are there written rules for combining the results of multiple observations? Yes __ No __

8. Are teachers required to sign evaluation forms? Yes __ No __

9. Is there a formal grievance procedure for teachers to follow in contesting negative evaluations? Yes __ No __

10. Is there a formal grievance procedure for teachers to follow in contesting violations of procedures? Yes __ No __

11. Is a remediation plan for teachers who receive negative evaluations included in the evaluation process? Yes __ No __

12. If teaching deficiencies are detected, is the evaluator obligated to develop a professional improvement plan for the teacher? Yes __ No __

13. Are there specific consequences for evaluators who violate procedures? Yes __ No __

IV. OPINIONS

This section contains questions concerning your opinions about the teacher evaluation system used in your school district. Please check only one response for each question.

1. How adequate is the teacher evaluation system in accomplishing the purposes for which it was designed?

More than adequate	Adequate	Less than adequate
_____	_____	_____

2. To what extent has the teacher evaluation system impacted the manner in which evaluators view classroom teaching?

Much impact	Some impact	Little impact
_____	_____	_____

3. To what extent has the teacher evaluation system served to communicate to teachers clearer expectations of classroom teaching?

Much impact	Some impact	Little impact
_____	_____	_____

4. To what extent has the teacher evaluation system served to improve instruction?

Much improvement	Some improvement	Little improvement
_____	_____	_____

5. To what extent has the teacher evaluation system led to better/more systematic supervision of teaching in your school district?

Much impact	Some impact	Little impact
_____	_____	_____